Speech Processing 2008/09

3rd Test

June 8th 2009

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)
   (a) Diagonal matrices are often adopted in continuous density HMM recognition systems in order to restrict the number of parameters to train.
   (b) A trigram-based language model generally leads to a higher perplexity than one based on bigrams.
   (c) Endpoint detection algorithms frequently involve several duration and energy thresholds.
   (d) A recognition system trained for the meetings domain will have a higher OOV rate when tested in the lecture domain.
   (e) The Viterbi beam search uses a time-asynchronous strategy.
   (f) VTLN methods can be used for the recognition of children voices.
   (g) Recognizers running in a forced recognition mode may be used to refine phone boundaries in a bootstrap process.
   (h) A recognition system trained for narrow-band telephone speech works equally well for wide-band speech.
   (i) A tree-based lexical structure is typical of small vocabularies.
   (j) In multi-pass decoding systems, pentagram language models are typically used in the second pass.
   (k) The study of the pronunciation variation at the level of the syllable led to the conclusion that the *disposable* part of the syllable is the onset.
   (l) Confidence values may be used to select the spoken material for adapting acoustic models in unsupervised speaker adaptation.
2. Increasing the number of Gaussian mixtures in continuous density HMM systems (tick all that apply)
   (a) depends on the availability of training data;
   (b) increases computational complexity;
   (c) is more important in speaker dependent systems than in speaker independent ones.
3. Give examples (one per class) of
   (a) Single-channel enhancement technique
   (b) Acoustic feature involving temporal processing
   (c) Smoothing technique that combines well trained and badly trained estimates for different orders of ngrams.
   (d) HMM training technique
4. Write in decreasing order of JND (Just Noticeable Difference)

(a) formant bandwithds
(b) F0
(c) formant frequencies

5. What is the typical frame size and spacing (ms) in HMM-based speech recognition?
6. Indicate an application domain for speech recognition with a highly confusable vocabulary.
7. The search or decoding problem tries to maximize a product of two probabilities. Which?
8. The performance of a phone recognizer (%PHONE ACCURACY) is typically
   (a) below 30%
   (b) between 30% and 50%
   (c) between 50% and 70%
   (d) between 70% and 90%
   (e) above 90%
9. The following extract was produced by a speaker independent large vocabulary continuous speech recognition system for the broadcast news show of 01-01-2009. This is the first story, characterized by a very loud background noise.
   Dois mil e nove começou numa explosão de fogo de artifício em praticamente todo o mundo.
   De olhos no céu milhões de pessoas visitaram as cores de dois mil e oito.
   E receberam dois mil e nove com tesouras de que não sejam bem assim.
   As previsões dos economistas e os políticos.
   Uma área de mais ambicioso espectáculo de Portugal de f um benefício da Madeira.
   Voltou em cheio nos hóteis de vida.
   Vitorino navios de cruzeiro, e no Brasil produções à baía do Funchal.
   The corresponding manual transcription is below:
   Dois mil e nove começou numa explosão de fogo de artifício em praticamente todo o mundo.
   De olhos no céu milhões de pessoas viraram as costas a dois mil e oito.
   E receberam dois mil e nove com desejos de que não sejam bem assim as previsões dos economistas e dos políticos.
   O mais ambicioso espectáculo de Portugal é o fogo de artifício da Madeira que voltou a encher os hóteis da ilha.
   E atraíu navios de cruzeiro e inúmeras embarcações à baía do Funchal.
   Ignoring punctuation and capitalization, compute the corresponding values of H (”correct”), D (”deletions”), S (”substitutions”), I (”insertions”), N (”total”), %Corr, %Acc and %WER. Compute as well the number of insertions of full stops.
10. Consider the training corpus that consists of the following sentences:
    O Luís gosta de ficção científica.
    A Madalena só gosta de filmes de terror.
    O Vasco não lê livros de ficção.
    O Luís não gosta de romances.
    A Madalena não vê filmes de ficção científica.
    Consider the test sentence:
    O Vasco gosta de livros de terror.
    (a) Compute the number of unigrams, bigrams and trigrams of the training corpus, and the dimension of the vocabulary.
    (b) Compute the probability of the test utterance using a bigram language model without any type of smoothing.
    (c) Compute the probability of the test utterance using a bigram language model with add-one smoothing.
    (d) Build a test sentence as long as possible with a non-zero probability according to a trigram model.
    (e) Write a quadrigram of the training corpus with more than one occurrence, or *none* if you cannot find it.

Test 3 - Answers

Name:	
Number:	

1. (3.6 val.) Indicate T or F:

a	b	c	d	e	f	g	h	i	j	k	l

2. (0.9 val.) Tick all that apply:

a	b	c

3. (2.4 val.)

a	
b	
c	
d	

4. (1.2 val.)

largest JND	
medium JND	
smallest JND	

5. (1.0 val.)

size	
spacing	

6 to 8 (1.0/1.0/1.0 val.)

6	
7	
8	

9. (3.5 val.)

H	D	S	I	N	% Corr	% Acc	% WER	Ins-Stop

10. (4.3 val.)

a)	
b)	
c)	
d)	
e)	